ATLANTIS PRESS
Advances in Computer Science Research (ACSR), volume 76
7th International Conference on Education, Management, Information and Mechanical Engineering (EMIM 2017)

Unmanned Boat Navigation Planning Based on Machine Learning Algorithm

Xiulai Li[1, a], Mingrui Chen[1, b*] and Lijie Yin[1, c]
[a]Hainan University, Haikou, Hainan, 570228, China
[a]lixiulai@hainu.edu.cn, [b]mrchen@hainu.edu.cn, [c]18315913978@163.com

Keywords: Unmanned boat; Q-learning; Unmanned technology; Algorithm

Abstract. This paper mainly introduces unmanned ship technology and the frontier technology of artificial intelligence, especially the Q-learning algorithm in machine learning. It mainly describes how the machine learning algorithm in artificial intelligence technology be applied in unmanned ship technology, so as to improve the efficiency and accuracy of unmanned ship.

Introduction

In recent years, unmanned vehicles are gradually familiar with the public. Google is the pioneer in the field of unmanned automotive research and development. April 16, 2016, Changan Automobile developed unmanned vehicles completed from Chongqing to Beijing, a full length of 2000 km highway test, although the gas station into the scene and other people need to meet, but in most cases still in the Unmanned state, which is also the longest distance in the country unmanned car experiment. In the unmanned field, compared to unmanned aerial vehicles, unmanned vehicles, unmanned ship development in a relatively backward state[1]. In recent years, with the internet of things, big data, cloud computing, artificial intelligence and other new ideas, new technology by leaps and bounds, ship automation level continues to improve, unmanned ship to achieve the technology support, no people driving a ship sailing in the world have the possibility of realization[2].

For unmanned sailing, the two most important technologies are automatic obstacle avoidance technology and route navigation technology[3]. Automatic obstacle avoidance technology is an important guarantee for the safe and stable driving of unmanned ships. For some areas of study, unmanned ships are required to sail according to the scheduled route. However, the size and direction of the ocean wave changes, requiring unmanned navigation direction and the engine's driving force also changes in order to travel in accordance with the scheduled route[4]. The early obstacle avoidance algorithm is based on the establishment of precise mathematical model of the controlled object and the detection of comprehensive environmental information. The control algorithm is complex. In general, the mathematical model of the controlled object is difficult to determine accurately and eventually lead to obstacle avoidance Not obvious. For unmanned navigation routes, straight lines or other scheduled routes are sailing in the waves, and the existing technology is small and the accuracy is not high. With the deepening of the study of machine learning theory, its application is also increasingly widespread, many excellent algorithms came into being. The machine learning algorithm introduced into the field of unmanned navigation technology, effectively improve the accuracy of unmanned navigation.

Machine Learning Overview

Mechanical Learning. Mechanical learning is memory learning, is the most primitive learning algorithm[5]. Mechanical learning is for each input of information and solve the problem into the knowledge base, when the problem again encountered, the direct access to the knowledge base, get the solution to the problem. This learning strategy does not require any reasoning process, the external input knowledge representation and system internal representation exactly the same, without any processing and conversion. While machine learning looks relatively simple, machine learning can produce unpredictable results because the computer's storage capacity is very large, the

retrieval speed is very fast, and the memory is accurate without any errors. As shown in Figure 1.

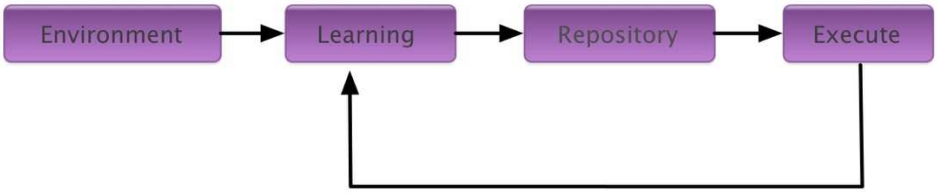

Figure 1. Machine learning

The following is a brief description of the machine learning algorithm. Let the problem be solved as follows: (p1, p2, ..., pn}, the problem is solved after the input information {x1, x2, ... xm}, so the record will be {{x1, x2, ..., xm}, {p1, p2, ..., pn}} into the knowledge base, and then when the problem {p1, p2, ..., pn} is encountered, query the knowledge base, remove {x1, x2, ..., xm} As a solution to the problem {p1, p2, ..., pn}. As shown in Table 1.

Table 1 Machine learning

Environment	X_1	X_2	...	X_n
Solving	P_1	P_2	...	P_n
Repository	(X_1, P_1)	(X_2, P_2)	...	(X_n, P_n)

As mentioned above, a system capable of implementing a mechanical learning algorithm requires only two basic skills, memory and retrieval. In addition, the reasonable arrangement of storage, the rational combination of information and the retrieval of the optimal direction of control is also the system should consider the issue[6]. The algorithm is easy to implement and computationally fast. However, because the system does not have the function of inductive reasoning, even if it is a similar problem for each different problem, it is necessary to have different records in the knowledge base, so it takes a lot of storage space, this is a typical space-time algorithm.

Q-Learning. An autonomous robot that can perceive its environment, how to choose the best action to achieve its goal by learning, which is the goal of Q-learning. The purpose of enhanced learning is to learn how to take optimal behavior in the corresponding observations by interacting with the environment[7]. Compared to mechanical learning, it has the following advantages, effectively solve the special circumstances of the environment. Second, you can put the whole system as a whole, which is more robust to some of these modules. Finally, Q-learning can be easier to learn a range of behaviors. When the robot makes every action in its environment, the instructor will raise the relevant reward or penalty information to indicate whether the current state is correct. Machine in the learning process, the environment provides input and output pairs, learning the task is to find a function to meet the relationship between these input and output pairs. In the Q-learning, the learning system adjusts the parameters of the system according to the state of the feedback signal from the environment[8]. The automaton responds to the environment, responds to the environment, receives an assessment of the environment, receives a new stimulus from the automaton, and automatically adjusts the learning system according to the last response of the automaton and the current input Parameters, Q-learning mechanisms such as shown in Figure 2.

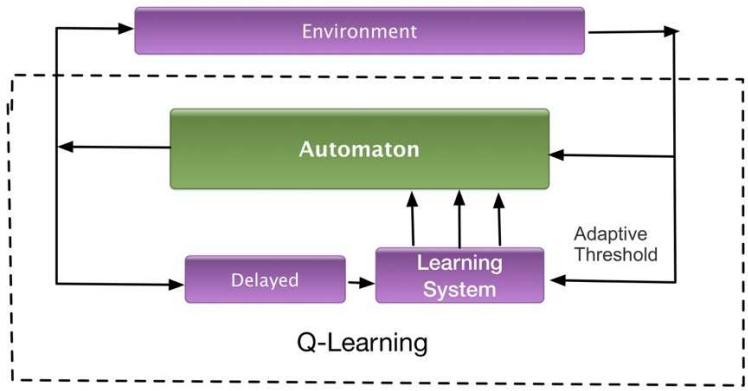

Figure 2. Q-learning mechanisms

Q-learning algorithm in the deterministic return and action assumptions under the Q-learning algorithm: (s said state, g said the action, Q (s, g) on the state s under the action of g to get an estimate of the overall return, r for this action Of the immediate return, γ for the discount factor, where 0 ≤γ<1)

(1) For each s, g initialize entry $Q(s,g) = 0$
(2) Observe the current state s
(3) Has been done repeatedly:
(4) Select an action g and execute it, which is the largest g for Q (s, g).
(5) Receive an immediate return r.
(6) Observe the new state s'.
(7) Update the entry for $Q(s',g')$ as follows:
(8) $Q(s,g) = r(s,g) + \gamma * \max Q(s',g')$.
(9) $s = s'$.

The Application of Machine Learning in Unmanned Boat Sailing

Unmanned Boat Simple Environmental Navigation Strategy. For example, unmanned ships in the course of the voyage, the default route is to ensure the obstruction of the premise of a straight line navigation. During the course of the voyage, the hull is changed by the intensity and direction of the waves and is unpredictable. It is clear for the unmanned boat itself. Therefore, unmanned ships in the process of navigation, continue to train the perception of the surrounding environment, and make the appropriate strategy, if the results of the implementation of the strategy in line with the default route to be rewarded. As shown in Figure 3 and Figure 4.

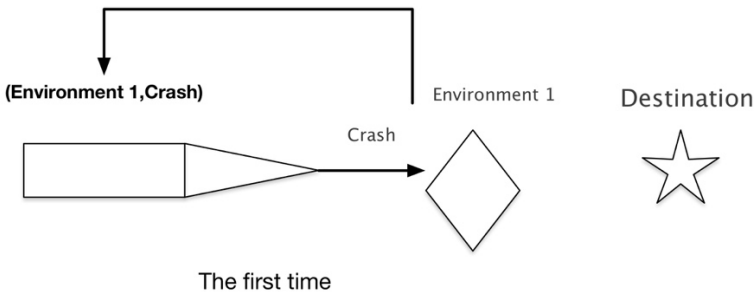

Figure 3. The first time lane

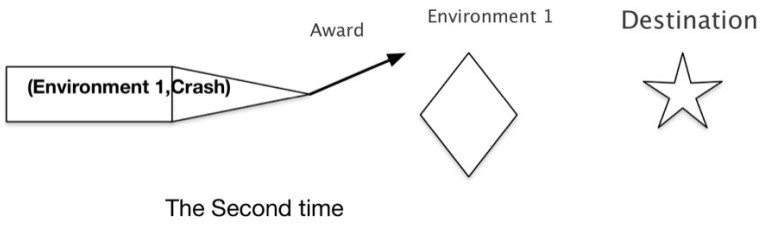

Figure 4. The second time lane

In every time t∈{ 0,1,2,…} in the unmanned aerial study, the algorithm and the environment interact with each other by executing the behavior **gt**, and you can get the observation **st** and the reward **ft**. In general, we assume that the environment is present in the Markovian nature, that is, the change of the environment can be described by the state transition probability as Eqs. (1).

$$Pass' = Pf\{st+1 = s'|st = s, gt = g\} \quad (1)$$

The next moment of the environment is observed only in relation to the observation and behavior of the moment, and it is not relevant to the observation and behavior of all the time before. And the expectation of the reward returned by the environment at time t+1 can be expressed as: Ras=E {ft+1|st=s, gt=g}. The enhancement learning algorithm performs the behavior at every moment The strategy can be expressed by the probability as Eqs. (2).

$$\pi(s, g, \theta) = Pf\{gt = a \mid st = s; \theta\}$$
(2)

Where θ is the policy parameter that needs to be learned. The best strategy is as Eqs. (3) and Eqs. (4).

$$Q*(s, g) = Es'[f+ \gamma maxg'Q*(s, a) \mid s, g] \quad (3)$$

$$L(w) = E[(f + \gamma \max g'Q*(s', g'; w) - Q(s, g; w))2] \quad (4)$$

Where γ is the discount factor in the enhanced learning and is used to represent the bonus discount obtained at a later time. The function Qπ(s, g) represents the reward that can be obtained using the strategy π selection behavior at the time after the state s and execution behavior g. We need to learn the best strategy to Q-learning, that is, to learn to get the highest reward strategy, through the exploration of the environment to learn the best strategy function π maximized fho(π).

Unmanned Boat Complex Environmental Navigation Strategy. Unmanned boat sailing, we only need to control the direction and speed of the ship so that the ship along the straight line or the default route, and avoid other obstacles can get the best strategy. But for more complex decision-making scenarios, we can not get the best strategy through short-term rewards. For example, games have two players, represented by white and sunspots, respectively. The intersection of the chessboard center line is where you can go. Two players are under the board under the white and sunspots. Once a white child or sunspot is surrounded by the opposite color of the child, then the film will be lifted, re-become a blank area. At the end of the game, all the blank areas are occupied or surrounded[9]. Occupation and surrounded by the larger side of the party to win.

In the chess strategy, we get from the environment observation st is the state of the board, that is, the distribution of white and sunspots. The act we perform is the position of the white or sunspot. And we finally get the reward can be based on whether the game win, win +1, the failure of the party -1. The game process can be represented by the following search tree: Each node in the search tree corresponds to a checkerboard state, with each side corresponding to a possible behavior. In the search tree shown in Fig 5, Black precedes the tree's root node, which corresponds to the initial state s0 of the board. G1 and g2 correspond to two possible positions of black (the actual go, the possible behavior is much more than two). Each action gi corresponds to the state si1 of a new board. Then the white go, White also has two kinds of walking b1 and b2, for each state of the board si1, two different ways to generate two different states[10]. So back and forth, until the end of the game, we can get the end of the game at the end of the game Black won the reward. We can get the best of these rewards.

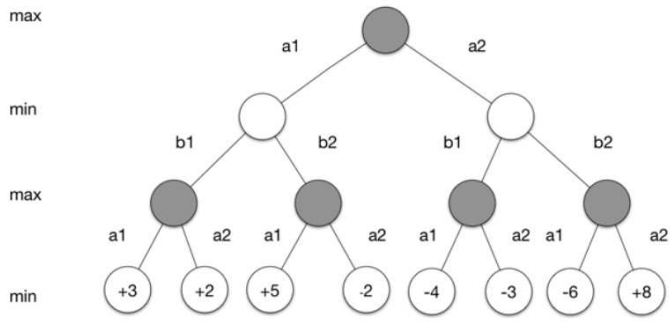

Figure 5. Search tree for chess strategy

Through this search tree, if given the black and white strategy $\pi = [\pi 1, \pi 2]$, we can define the black value of the function for the blacks on both sides when executing the strategies $\pi 1$ and $\pi 2$ respectively, and finally Black can get the reward The expectation as Eqs. (5) and Eqs. (6).

$$v\pi(s) = E\pi[Gt|St = s] \quad (5)$$

$$\nabla w = \alpha(v(st + 1; w) - v(st; w))\nabla wv(st; w) \quad (6)$$

Black needs to find the best strategy that needs to be optimized for the worst case where Black can get the reward. We define this value function as the minimum maximum function. Black's best strategy is to be able to achieve this value function strategy $\pi 1$.

Unmanned boat sailing uses the state of the environment as input, using machine learning algorithms as learning algorithms. The environmental reward is defined as the distance traveled straight in the unit, and if the unmanned boat is out of a straight line or collided, an additional penalty will be given. The state of the environment includes the speed of the unmanned boat, the acceleration, the height of the water at the side of the unmanned hull, and the distance of the bow, the distance of the nearest obstacle in each direction, and so on[11]. Unmanned boat behavior, including forward gear, back gear, acceleration, deceleration, to the left to play the steering wheel, right to play the steering wheel and so on. Through the machine to learn the approximate minimum maximum function, according to the state of the environment at the last moment to get the reward.

Summary

Artificial intelligence is the hot topic of the current computer industry research, machine learning is the focus of artificial intelligence, for a lot of industry research has a wide range of applications, the effective introduction of new technologies into the existing industry, is conducive to the development of science and technology. Unmanned aerial vehicles in the field of marine research use more and more widely, effectively improve the efficiency of unmanned ships will promote marine research, is conducive to China's development strategy. Unmanned technology in various areas to gradually promote the development of unmanned intelligent control in the future will likely have a broader prospect.

References

[1] Y. Liu and N. Noguchi: Engineering in Agriculture, Environment and Food, Vol. 9(2016), No. 1, p.21.
[2] C. Sonneburg: Modeling Identification and Control of an Unmanned Surface Vessel, Vol. 2(2012), No. 4, p.30.
[3] L. McNinch and H. Ashrafiuon, *Proceedings of American Control Conference*, (San Francisco, USA, July 1–3, 2011). Vol. 1, p.184.
[4] D. Chwa, *IEEE Transactions on Control Systems Technology*, Vol. 19(2011), No. 6, p.1357.
[5] Zhiqiong Wang, Junchang Xin: Tsinghua Science and Technology, Vol.22(2017), No.1, p.160.
[6] Soroosh Golbabaei, Amirali Vahid: *2016 23rd Iranian Conference on Biomedical Engineering*

and 2016 1st International Iranian Conference on Biomedical Engineering, (Iran, November 24–25, 2016). Vol.1, p.35.
[7] Sitt Sidehabi, Indrabayu Statistical: *2016 International Conference on Computational Intelligence and Cybernetics*, (South Sulawesi, Indonesia, November 22,2016), Vol.1, p63.
[8] Kevin P. Murphy: *Machine Learning* (The MIT Press, USA,2012).
[9] Ian Goodfellow: *Deep Learning* (The MIT Press, USA,2016).
[10]Christopher Bishop: *Pattern Recognition And Machine Learning*(Springer, Germany,2007).
[11]Y. Kaizu, M. Iio and H. Yamada: *Biosystems Engineering*, Vol.109(2011), No. 4, p. 338.